

基于大语言模型的电商金融分析系统

计金（双） 刘宇晨 20002515

摘要：随着电商金融的快速发展，个性化推荐系统成为提升用户体验和增加用户粘性的关键技术。为了更好地理解用户的需求和偏好，本文设计了一个名为 Seq2Label 的大语言模型，旨在将用户评论精简为商品标签。该模型采用了双向 GRU 和 Luong 注意力机制等深度学习框架，再通过 Kmeans 聚类算法，实现了一个能够不断自我更新的推荐系统。系统通过分析和处理源源不断的用户评论，自动转换成 N 个最能代表商品特点的标签。这些标签不仅反映了用户的真实诉求，而且通过 Fine-tuning 模型和词嵌入技术的结合，能够为用户推荐最适合他们的商品。不仅如此，通过分析用户的购买行为和个人财务信息，本系统为用户提供贷款服务，有效增加用户粘性和平台吸引力。

关键词：双向 GRU、Luong 注意力机制、Fine-Tuning、词嵌入技术、电商金融

1 研究背景

1.1 选题背景

近年来，伴随着深度学习与大语言模型的飞速发展，我们见证了自然语言处理（NLP）领域的一次又一次突破。自从 Transformer 模型问世以来^[1]，在大语言模型领域涌现出了一系列高度可用的模型，包括 GPT^[2]系列和 BERT^[3]及其变体等。这些模型在自然语言处理任务中取得了令人瞩目的成果，并推动了该领域的创新发展。这些进步不仅推动了机器学习理论的深入发展，也为实际应用提供了前所未有的可能性。特别是在文本生成、情感分析、机器翻译、语言理解等方面，大语言模型展现出了惊人的能力，极大地丰富了人们对语言智能化处理的认识和期待。

随着模型性能的不断提升，其应用范围也在不断扩大。在电子商务、金融科技、股市情感预测、社交媒体分析等多个领域，大语言模型已成为解决关键业务问题、提升服务质量、增强用户体验的重要工具^[4]。尤其是在电子商务领域，通过深入分析用户评论、反馈和行为数据，大语言模型能够帮助企业更好地理解用户需求，实现更加个性化的商品推荐，从而提升用户满意度和购物体验。

1.2 研究意义

在当前的电子商务领域，推荐系统的发展已达到了一个高度成熟的阶段，能够利用用户的购买历史、浏览习惯等数据，通过复杂的算法为用户推荐他们可能感兴趣的商品^[5]。这种基于历史数据的推荐技术，虽然在很大程度上优化了用户的购物体验，增加了平台的销售额，但仍存在一些不可忽视的局限性。特别是当用户面对具有模糊性或多样

化需求的购物决策时，现有的推荐系统往往难以准确捕捉到用户的真实意图，导致推荐结果可能与用户期望存在偏差^[6]。

随着电子商务与金融科技的融合发展，用户的购物行为和消费需求变得更加多元和复杂。用户不仅仅在寻找特定的商品，更希望通过购物过程获得满足个性化需求的体验。在这样的背景下，传统推荐系统的局限性就显得尤为突出。因此，探索新的推荐技术，以更准确地理解和满足用户的个性化需求，成为了电商平台提升竞争力的关键。

本文整体的框架，一方面通过引入生成式大语言模型和自己构建 Seq2Label 模型来优化推荐系统，可以接受用户以自然语言表达的购买诉求，通过智能分析和处理这些文本数据，生成与用户需求高度相关的商品特征标签。这些标签不仅细化了用户的需求，也为推荐系统提供了更为丰富和精确的输入，从而大大提高了推荐的准确性和个性化程度；另一方面通过分析用户的购买行为和个人财务信息，为用户提供贷款服务，大大拓宽了电商金融分析系统的应用范围，并增强了其对用户需求的响应能力。

贷款服务的引入可以有效增加用户粘性和平台的吸引力。通过提供贴合用户财务需求的贷款选项，平台能够在用户的购买决策过程中起到关键作用，进一步促进销售和用户满意度的提升。其次，这种服务能够为电商平台带来更加深入和全面的数据分析。通过收集和分析用户的财务信息和购买行为，平台可以更准确地识别用户需求，预测市场趋势，并据此优化商品推荐算法，提升个性化服务的水平。

综上所述，本文通过引入和应用最新的大语言模型技术和智能借贷功能，为电商金融领域的发展提供了新的视角和方法论，不仅在学术上具有创新性，也在实践中具有广泛的应用前景和商业价值。通过进一步的研究和实践，有望推动电子商务平台在满足用户个性化需求、提升用户体验和拓展业务范围等方面取得新的进展。

1.3 可持续发展的意义

本文研究的利用生成式大语言模型和智能借贷功能优化电商推荐系统，在可持续发展方面具有重要的意义，同时也充分考虑到环境保护和可持续发展的需要。

1.3.1 环境可持续性

通过高效的推荐系统减少了不必要的购买和退货，有效降低了物流运输过程中的碳排放和包装材料的浪费。此外，智能推荐系统能够鼓励消费者购买环保产品，从而促进环境保护和生态平衡。

1.3.2 经济可持续性

通过个性化推荐和贷款服务，电商平台能够更好地满足消费者需求，提升用户体验，从而增加用户粘性和销售额。这不仅促进了电商平台的经济效益，也通过促进消费而支持了地方和全球经济的发展。

1.3.3 社会可持续性

智能推荐系统通过提供贴合用户需求的产品和服务，增强了消费者的满意度和生活质量。同时，智能借贷功能增加了金融服务的覆盖范围，提高了金融包容性，使得更多的人能够享受到便捷的金融服务。

1.3.4 优化能源使用

根据 Strubell 等人（2019）的研究，深度学习模型在自然语言处理（NLP）中的应用虽然取得了显著进展，但同时也伴随着巨大的能源消耗^[7]。本文提出的自研 Seq2Label 模型在保持推荐系统效率和准确性的同时，大大降低了计算量和能耗。这一优化表明，相较于传统的通用大语言模型如 ChatGPT，在处理大量数据时，我们的模型能更有效地减少能源消耗。这不仅体现了科技创新在提高计算效率方面的潜力，也强调了在环境保护领域采取积极行动的重要性。

2 文献综述

本文的推荐系统流水线结合了多种不同领域的大语言模型框架以及机器学习框架。

2.1 词向量嵌入技术

词向量嵌入技术是自然语言处理（NLP）领域的一项核心技术，它将词语转换为计算机可以理解的数学向量^[8]。通过这种方式，模型能够捕捉到词语之间的语义关系，包括相似性和上下文关联性。词向量的引入极大地提升了机器对自然语言的理解能力，是众多 NLP 任务（如文本分类、情感分析、机器翻译等）的基础。

分布式表示（也称为词嵌入）技术能够将词表示为稠密的向量，这些向量在低维空间中。这种表示方法的基本假设是“语境相似的词其语义也相似”，这一假设也被称为分布假说。主要的词嵌入模型包括 Word2Vec、GloVe 和 FastText 等。

2.1.1 Word2Vec

Word2Vec 是由 Google 团队提出的模型^[9]，有两种架构：连续词袋模型（CBOW）和跳字模型（Skip-gram）。CBOW 预测目标词基于上下文，而 Skip-gram 则反其道而行之，预测上下文基于目标词。Word2Vec 的目标是使得模型学习到的词向量能够将具有相似上下文的词聚集在向量空间中的相近位置。本文的 Seq2label 模型采用了 Word2Vec 模型作为其底层词嵌入层。

2.1.2 GloVe

GloVe（Global Vectors for Word Representation）是一种在自然语言处理（NLP）领域广泛使用的词嵌入技术。该方法由 Pennington 等人提出^[10]，旨在通过捕捉单词之间的全局共现统计信息来有效地学习词向量表示。与传统的基于局部上下文信息的模型不同，

GloVe 构建一个大型的共现矩阵，描述单词之间在整个数据集中的共现关系，并通过矩阵分解技术学习词向量。这样不仅可以捕捉到丰富的语义和语法关系，还能有效地利用统计信息，提高词向量的质量和表现力。GloVe 方法因其结合了全局统计信息和局部上下文特征的优点，已被广泛应用于各种 NLP 任务中，如文本分类、情感分析和机器翻译等^[11]。

2.1.3 FastText

FastText 是 Facebook 研究团队提出的一种词嵌入方法^[12]，与 Word2Vec 类似，但它在训练过程中不仅考虑词本身，还考虑词内的子词（subword）。这种方法对处理形态丰富的语言（如德语）特别有效，因为它能够捕捉到词缀等语言特性，提升模型对未见词的处理能力。

2.1.4 text-embedding-ada-002

"text-embedding-ada-002" 是 OpenAI 推出的一款革命性的文本嵌入模型，以其高性能和成本效益著称，旨在为文本搜索、文本相似度和代码搜索任务提供优化解决方案。它通过集成并取代五个不同的模型，不仅在大多数任务中超越了先前的顶尖模型达芬奇，而且其成本仅为达芬奇的 0.2%。text-embedding-ada-002 的特色在于其能力强大，无论是在文本搜索、代码搜索还是句子相似性任务上，它均展现出卓越的性能，并在文本分类等领域达到了与旧模型相媲美的表现。此模型不仅精通理解文本的深层含义，还能将复杂的文本信息高效压缩成数值向量，支持广泛的自然语言处理任务。作为基于深度学习的预训练模型，text-embedding-ada-002 已在庞大的文本数据集上进行过训练，为用户直接使用提供便利。

2.2 Seq2Seq 模型架构

序列到序列(Seq2Seq)模型架构自 2014 年由 Sutskever 等人以及 Cho 等人在其论文中提出以来^[13]，已成为处理诸如机器翻译、文本摘要、问答系统等自然语言处理任务的基石。该模型架构主要由两部分组成：编码器(encoder)和解码器(decoder)，通过这两个组件将输入序列转换为输出序列。Seq2Seq 模型的关键在于能够处理不等长的输入和输出序列，为复杂语言处理任务提供了有效的解决方案。

2.2.1 门控循环单元（GRU）

门控循环单元（Gated Recurrent Unit）是为了解决传统递归神经网络（RNN）在处理长序列数据时常遇到的梯度消失问题而设计的先进结构。GRU 通过引入两种特殊的门结构——更新门和重置门——优化了信息的流动过程，提高了模型对长期依赖关系的捕捉能力。更新门的作用是帮助模型决定在当前状态下应该保留多少之前的信息，而重置门则决定放弃多少历史信息，以便更有效地处理当前输入信息。GRU 因其计算效率较高且在多种序列处理任务中表现出色，而广泛应用于自然语言处理（NLP）领域。

Sutskever 等人的工作进一步展示了使用神经网络进行序列到序列学习的强大潜力，其中 GRU 的使用对于提高模型性能尤为关键^[14]。这些研究表明，GRU 不仅解决了梯度消失问题，还提供了一种比长短时记忆网络（LSTM）更为高效的选择，使其成为处理复杂序列任务的重要工具。

2.2.2 Luong 注意力机制

Luong 注意力机制，由 Luong 等人在 2015 年提出^[15]，是另一项提高 Seq2Seq 模型性能的关键技术。不同于早期的注意力机制，Luong 提出了几种计算注意力权重的方法，包括点积（dot）和一般（general）方法。Luong 注意力机制允许解码器在生成每个词时更加专注于编码器输出的相关部分，从而提高了翻译质量和相关文本生成任务的性能。

2.2.3 编码器-解码器架构

Seq2Seq 模型的核心是其编码器-解码器架构。编码器部分负责处理输入序列，将其转换为固定长度的向量表示，这一表示捕获了输入序列的主要信息。解码器则负责将这个向量表示转换为输出序列。这种架构的优势在于其灵活性和通用性，能够应对各种长度的输入和输出序列。通过集成 GRU 和 Luong 注意力机制，Seq2Seq 模型能够更有效地处理序列间的复杂依赖关系，提高了处理自然语言的能力和效率。

2.3 模型微调

模型微调（Fine-tuning）是迁移学习技术的一个关键应用，它依托于预训练模型，并在此基础上通过额外训练来适应特定的下游任务^[16]。这一过程特别适用于自然语言处理（NLP）领域，因为预训练模型，如 OpenAI 的 GPT 系列，已经在广泛的数据集上学习到了丰富的语言知识和模式^[17]。利用这些模型，开发者可以通过相对较少的特定任务数据来实现快速优化，无论是进行文本分类、情感分析还是构建问答系统。

OpenAI 的微调方法让开发者可以通过提交特定格式的数据文件来定制化其模型，比如 ChatGPT。这通常涉及到使用 JSONL 格式的文件，每行包含一个独立的 JSON 对象，使得模型能够以流式的形式处理大量数据。在微调 ChatGPT 模型时，每个 JSON 对象一般由两部分构成：一部分是输入文本（模拟用户的问题），另一部分是期望的输出文本^[18]。

通过为模型提供包含特定任务示例对话的 JSONL 文件，模型不仅学习语言的基本结构，也学习如何为特定任务生成合适的回答或内容。微调期间，模型参数会根据这些数据进行调整，目的是减少预测输出与实际输出之间的差异，使模型更适应该任务。这种细微调整显著提升了模型在特定任务上的性能，特别是在预训练模型未能充分覆盖的领域或特定情境中。

微调的优势在于能够借助预训练模型强大的通用性能力，并通过少量特定领域的数据进行调整，使其更适合特定的应用场景。这种方法节省了从头开始训练模型所需的大

量时间和资源，同时保持了模型的高性能和灵活性。

2.4 文本聚类

文本聚类是自然语言处理（NLP）领域中的一个重要任务，它旨在自动将文本集合分组，使得同一组内的文本在内容上更加相似，而不同组之间的文本则相对不同。这一过程不依赖于预先标注的类别信息，是一种无监督学习方法。文本聚类的应用范围广泛，包括信息检索、文档管理、话题发现和内容推荐等。

2.4.1 K-means 聚类

K-means 聚类是一种广泛使用的聚类算法，最初由 MacQueen (1967)提出^[19]。该算法通过迭代优化，将数据集分成 K 个预定义数量的不相交子集（聚类），以使得聚类内部的数据点尽可能相似，而不同聚类之间的数据点尽可能不同。K-means 聚类的关键在于选择合适的 K 值以及初始的聚类中心。MacQueen 的工作为后续的聚类算法研究奠定了基础，特别是在处理大规模数据集时的应用^[20]。

在文本聚类领域，K-means 算法也显示出了其有效性。Zhang 等人在 1996 年提出的 BIRCH 算法^[21]，即一种针对大型数据库的高效数据聚类方法，实际上是在 K-means 的基础上进行的改进。BIRCH 算法通过构建一个称为 CF Tree（Clustering Feature Tree）的树状结构来高效地处理大量数据，有效地提升了 K-means 在处理大规模文本数据时的性能和可扩展性。

2.4.2 层次聚类

层次聚类是另一种常见的聚类方法，它不同于 K-means 的划分聚类方式，层次聚类通过构建一个多级的聚类结构，这个结构可以是聚合的（自底向上）或者是分裂的（自顶向下）。Ester 等人在 1996 年提出的 DBSCAN 算法，是一种基于密度的空间聚类算法，该算法能够发现具有噪声的大型空间数据库中的任意形状的簇，从而克服了 K-means 聚类仅能发现凸形簇的限制。DBSCAN 通过判定核心点以及可达性的概念，自动地确定聚类的个数，强化了对数据空间分布特性的利用，是层次聚类研究中的一个重要里程碑^[22]。

在整合这些论文内容到你的文献综述时，可以强调 K-means 和层次聚类算法在处理文本数据，尤其是在自然语言处理领域中的重要性和应用。此外，提及这些算法如何随着时间的推移而被改进和优化，以及它们在实际应用中面临的挑战和解决方案，将有助于展示聚类技术在文本处理中的进展和价值。

2.5 随机森林算法

随机森林算法由 Breiman (2001) 提出^[23]，是一种利用多个决策树的预测结果进行汇总以提升整体预测性能的集成学习技术。该算法通过对原始数据集进行自助采样来生成多个不同的数据子集，每个子集用来训练一个决策树。在决策树的构建过程中，随机森

林会在分割节点时随机选择一部分特征，这种方法不仅增加了模型的多样性，还有效降低了过拟合的风险。通过汇总所有决策树的预测结果，随机森林最终输出预测结果，这种策略显著提高了预测的准确性和稳定性。

在二分类任务中，随机森林算法特别适用于处理数据不平衡的情况。Chen, Liaw, 和 Breiman (2004) 探讨了随机森林在处理不平衡数据时的有效性，指出通过适当的样本采样和权重调整，随机森林能够有效改善少数类的预测性能[24]。在这种情况下，随机森林通过构建多个决策树，并利用多数投票原则来确定最终的分类结果，这不仅提高了模型对少数类别的识别能力，也保持了整体预测性能的高度。这使得随机森林在多个领域的二分类问题中，如欺诈检测、疾病诊断和客户流失预测等，成为一个非常强大和受欢迎的工具。

3 技术路线

3.1 开发工具与环境

本项目的技术栈涵盖了现代 Web 开发和机器学习领域的先进技术，具体使用了 Gin 框架和 Vue Vite 框架分别开发电商金融系统的后端和前端。在机器学习模块，采用 Pytorch 和 Transformer 框架进行模型的训练，同时使用 Sklearn 库中的 K-means 算法对结果进行聚类处理。数据存储方面，结构化数据和词向量数据分别存储在 MySQL 和 PostgreSQL 数据库中。此外，本项目还利用了 OpenAI 平台提供的 gpt-3.5turbo 接口进行模型的 fine-tuning 训练以及词向量嵌入，以增强系统的智能化处理能力。

3.1.1 Gin 框架

Gin 是一个用 Go 语言编写的高性能 Web 框架。它提供了一套简洁有效的 API，使得开发者能够快速构建高性能的 Web 应用。Gin 框架的设计哲学是提供最小的、最快的框架，而不牺牲灵活性。它通过使用 httprouter 路由库来实现路由功能，该库优化了路由的匹配过程，从而使 Gin 成为处理 Web 请求的高效框架。

3.1.2 Vue Vite 框架

Vue Vite 是一个现代化的前端开发框架，结合了 Vue.js 框架和 Vite 构建工具的优势。Vue.js 是一个渐进式 JavaScript 框架，用于构建用户界面，特别是单页应用（SPA）。Vite 作为一个前端构建工具，提供了快速的冷启动、即时模块热更新（HMR）和真正的按需编译。Vue Vite 框架通过利用现代浏览器支持的 ES 模块导入特性，实现了无需打包操作的开发服务器，大大提高了开发效率。

3.1.3 PyTorch

PyTorch 是一个开源的机器学习库，广泛应用于计算机视觉和自然语言处理等领域。

它提供了强大的张量计算功能以及动态计算图的特性，使得模型的设计、实验和调试过程更为直观和灵活。PyTorch 的核心是提供了一个类似于 NumPy 的张量库，但它支持 GPU 加速，可以显著提高计算效率。

3.1.4 PostgreSQL

PostgreSQL 是一个功能强大的开源对象关系数据库系统，以其可靠性、健壮性和性能而闻名。本文利用 pgvector 插件，旨在为数据库提供高效的向量搜索能力。它允许用户在 PostgreSQL 中存储和查询向量数据，非常适合于实现词向量等机器学习应用的数据存储需求。pgvector 通过优化向量之间的距离计算，支持快速的相似性搜索操作，使其成为处理大规模向量数据的理想选择。

3.1.5 MySQL

MySQL 是一个广泛使用的开源关系型数据库管理系统，它是基于结构化查询语言的。由于其高性能、可靠性和易用性，让其成为了许多 Web 应用程序的首选数据库。MySQL 在存储后端系统的结构化数据方面具有明显优势，它提供了强大的数据存储和检索能力，能够有效地管理大量数据。其事务支持、并发控制和复杂查询优化等特性，使得 MySQL 在数据一致性和访问效率方面表现出色。此外，MySQL 的可扩展性和灵活性也为用户在数据增长和应用程序扩展时提供了便利，允许用户根据需求调整数据库架构，而不会牺牲性能。

3.1.6 Docker

Docker 是一个开源的容器化平台，它允许开发者将应用及其依赖打包成一个轻量级、可移植的容器，然后在任何支持 Docker 的系统上运行这个容器。这种技术提供了软件开发和部署的一种更加快速、可靠的方法。Docker 容器相对于传统的虚拟机更加轻量级，因为它们共享同一操作系统内核，而不是为每个应用程序虚拟化一个完整的操作系统。这意味着启动速度更快，资源占用更少。

3.2 数据集

3.2.1 电商数据集

本文基于 Jianmo Ni 所整理的“Amazon Review Data (2018)”数据集^[25]，特别关注其中的“Grocery and Gourmet Food”部分作为研究对象。该数据集是 2014 年 Amazon 评论数据集的更新版本，由加州大学圣迭戈分校的 Jianmo Ni 负责维护。本数据集包含了评论、产品元数据（描述、类别信息、价格、品牌和图像特征）。本文的电商系统的商品数据来自数据集中的产品元数据。

在本研究中，我们专注于数据集中的“Grocery and Gourmet Food”部分，旨在开发一个名为 Seq2label 的推荐系统。为了训练该系统，我们采用了 OpenAI 提供的 GPT-3.5

API，对选定的评论数据块进行处理，请求 GPT-3.5 生成参考答案作为模型的训练集。鉴于 OpenAI 接口的使用成本较高和响应速度的限制，我们仅对数据集的一小部分数据进行了这样的训练。

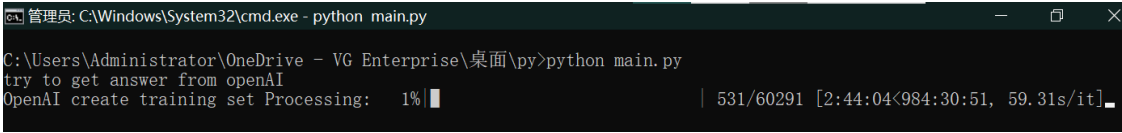


图 3-1 调用 OpenAi 接口获得训练集

通过这种方式，我们利用了大规模的用户评论数据和高级的自然语言处理技术，以期提高推荐系统的准确性和用户体验。此外，本研究的数据处理和模型训练方法也为其他领域的研究者提供了一个利用大规模文本数据和深度学习模型进行研究的实例。

3.2.2 贷款申请数据集

本项目旨在通过用户的购买记录和其它财务信息为用户提供贷款功能。其中通过用户购买历史记录判断贷款的所有数据集均来自 Kaggle 上的公开数据集¹²。

3.3 系统功能简介

在本研究中开发的电商金融系统中，前后端系统为用户提供了全面的购物体验，支持用户注册、浏览商品、添加商品到购物车、下单、查看订单以及获取个性化商品推荐等功能。管理端的设计使得商城管理员能够高效地编辑商品信息、商城公告以及配送信息，确保了商城运营的灵活性和高效性。

如图 3-2 所示，深度学习模块的核心是利用 Pytorch 训练的 Seq2label 的大语言模型，该模型基于 Transformer 架构，结合了 Kmeans 聚类算法，构建了一个能够持续自我更新的智能推荐系统。通过对大量用户评论的深入分析和处理，seq2label 模型能够自动提炼出能够最准确反映商品特性的标签，进而利用这些标签为用户推荐最合适的商品。这一过程不仅显著提升了推荐的准确性，也极大地丰富了用户的购物体验。模型通过细致的 fine-tuning 和词嵌入技术的应用，进一步增强了推荐系统的性能。

¹ <https://www.kaggle.com/datasets/bhavikjikadara/loan-status-prediction/data>

² <https://www.kaggle.com/datasets/mkechinov/ecommerce-purchase-history-from-electronics-store>

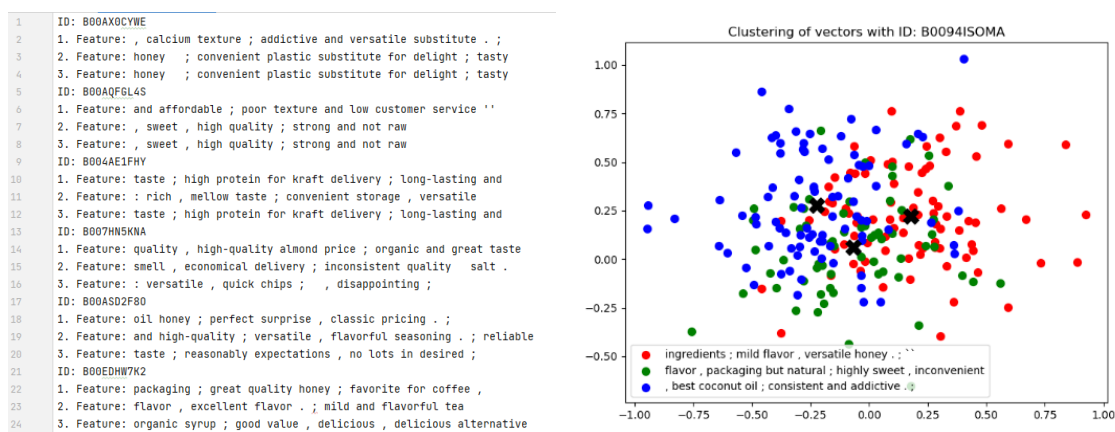


图 3-2 聚类后商品评论标签

如图 3-3 所示，系统通过整合用户的直接输入和行为反馈数据来生成推荐。用户输入首先经过定制微调的 GPT-3.5 模型处理，产生特征向量，同时用户行为反馈通过 MySQL 和 KMeans 聚类算法处理以发现数据模式。这些特征和聚类结果被转化为标签，再通过 PostgreSQL 进行向量搜索，最终系统根据这些信息推荐四个相关的商品或服务给用户。整个过程是一个结合了自然语言处理和机器学习技术的复杂推荐引擎，旨在提供个性化的用户体验。

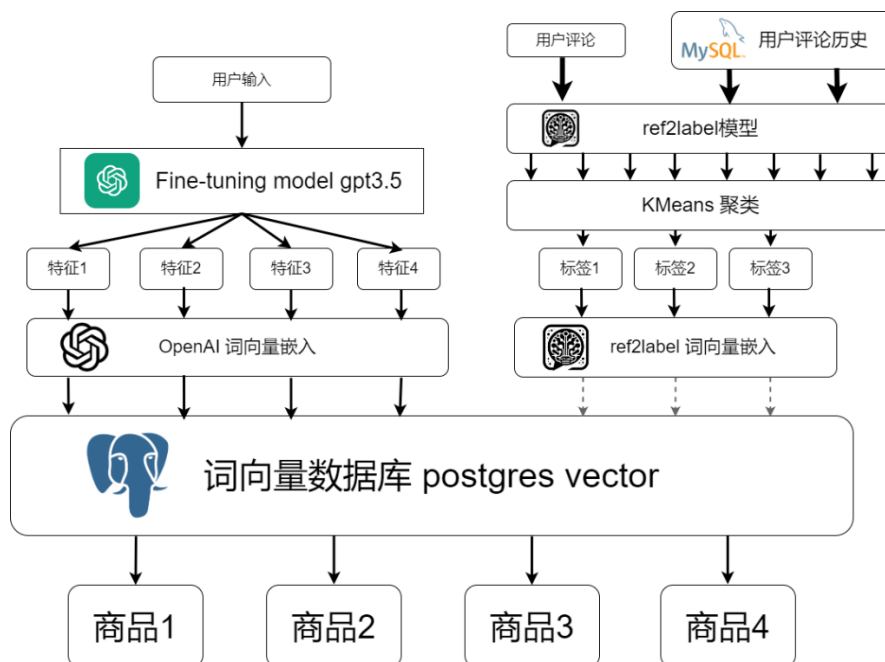


图 3-3 推荐系统流程图

3.3.1 用例图

如图 3-4，用户的用例图如下：

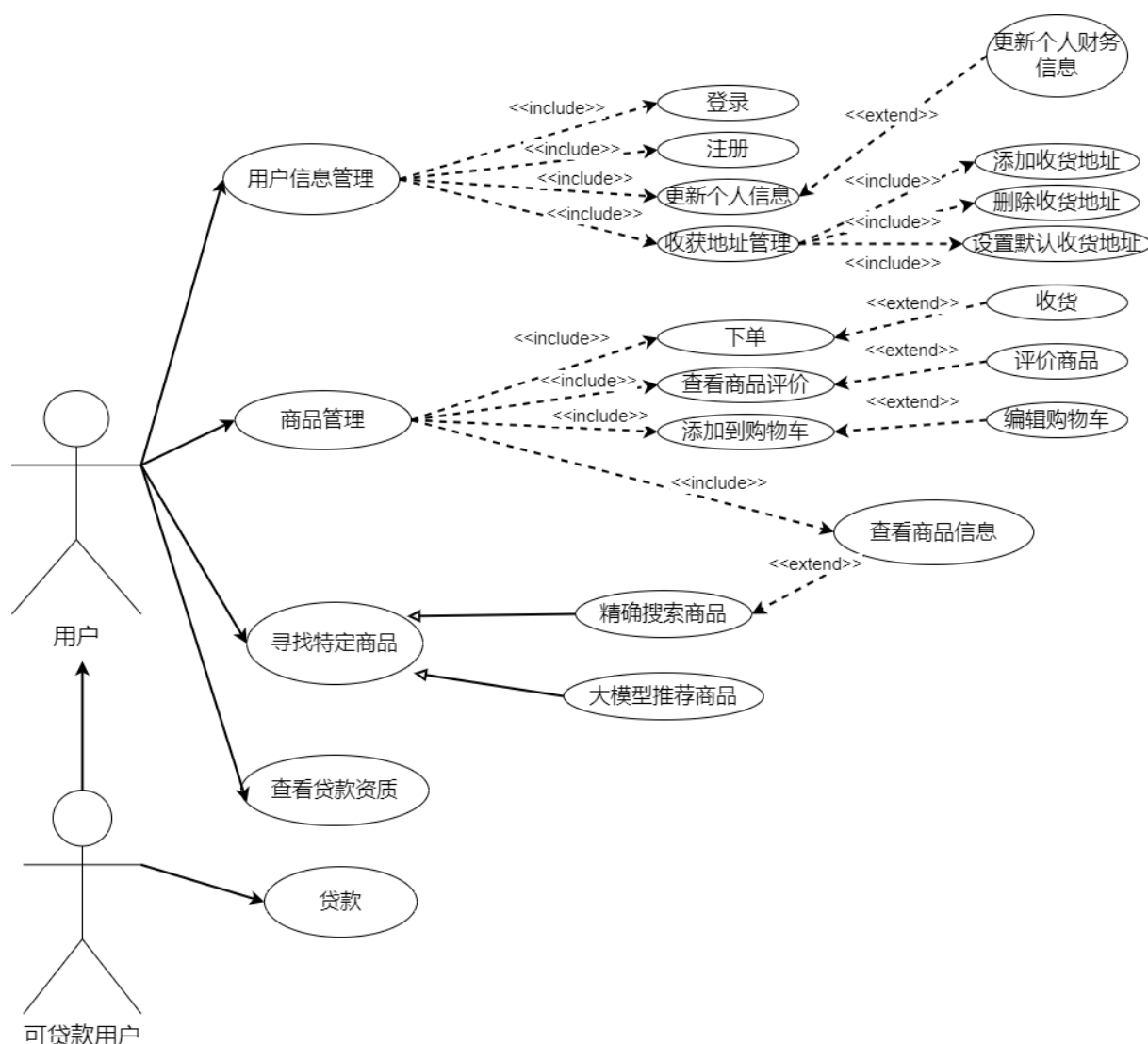


图 3-4 用户的用例图

1. 用户可以登录、注册系统：用户可以创建一个新的账户或者使用现有的账户信息登录系统。所有用户数据，包括个人信息、登录凭证等，都将安全地存储在服务器的数据库中。
2. 用户可以查看商品：注册用户可以浏览平台上的所有商品。这包括查看商品详情、价格、评价等信息。
3. 用户可以添加商品到购物车：用户可以选择心仪的商品加入到他们的购物车中，以便之后进行购买。
4. 用户可以下单：用户可以从购物车中选取商品，确认购买信息后，完成订单的创建。订单信息将被记录并用于后续的配送处理。
5. 用户可以获取推荐商品：根据用户的浏览历史和购买习惯，系统会自动推荐相关商品。这一过程可能涉及到深度学习或其他智能算法，以提高推荐的准确性。
6. 用户可以编辑个人信息：用户可以更新他们的个人信息，包括但不限于联系方式、配送地址等，确保订单能够正确配送到达。

如图 3-5，管理员的用例如下：

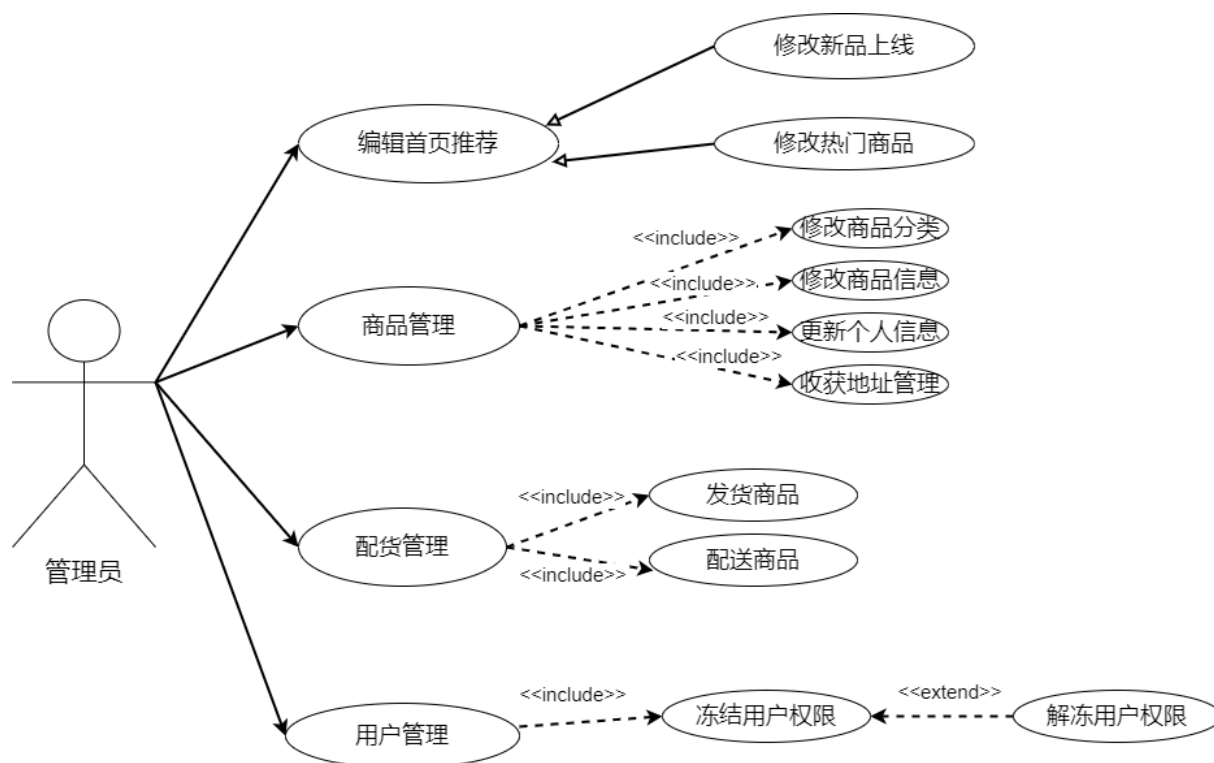


图 3-5 管理员的用例图

1. 管理员可以编辑商品：管理员有权限对商品信息进行编辑，包括更新商品描述、价格、库存量等信息。
2. 管理员可以编辑商品分类：管理员可以添加、修改或删除商品分类，以便更好地组织平台上的商品，帮助用户更容易地找到他们想要的商品。
3. 管理员可以对已有的订单进行配货：管理员负责处理订单，包括确认订单信息、准备货物、安排发货等，确保客户能够及时收到商品。
4. 管理员可以管理用户权限：管理员可以对用户的权限进行管理，包括限制或恢复用户的购买权限，确保平台的正常运营。
5. 管理员可以管理商城首页推荐：管理员可以选择特定的商品或促销信息展示在商城首页，吸引用户关注和增加销量。

3.3.2 系统功能模块简介

在本研究中开发的电商金融系统是一个综合性的电子商务解决方案，如图 3-6，它集成了前后端系统、管理端功能以及先进的深度学习技术，旨在为用户和管理员提供高效、个性化的服务。以下是系统的主要功能模块简介。

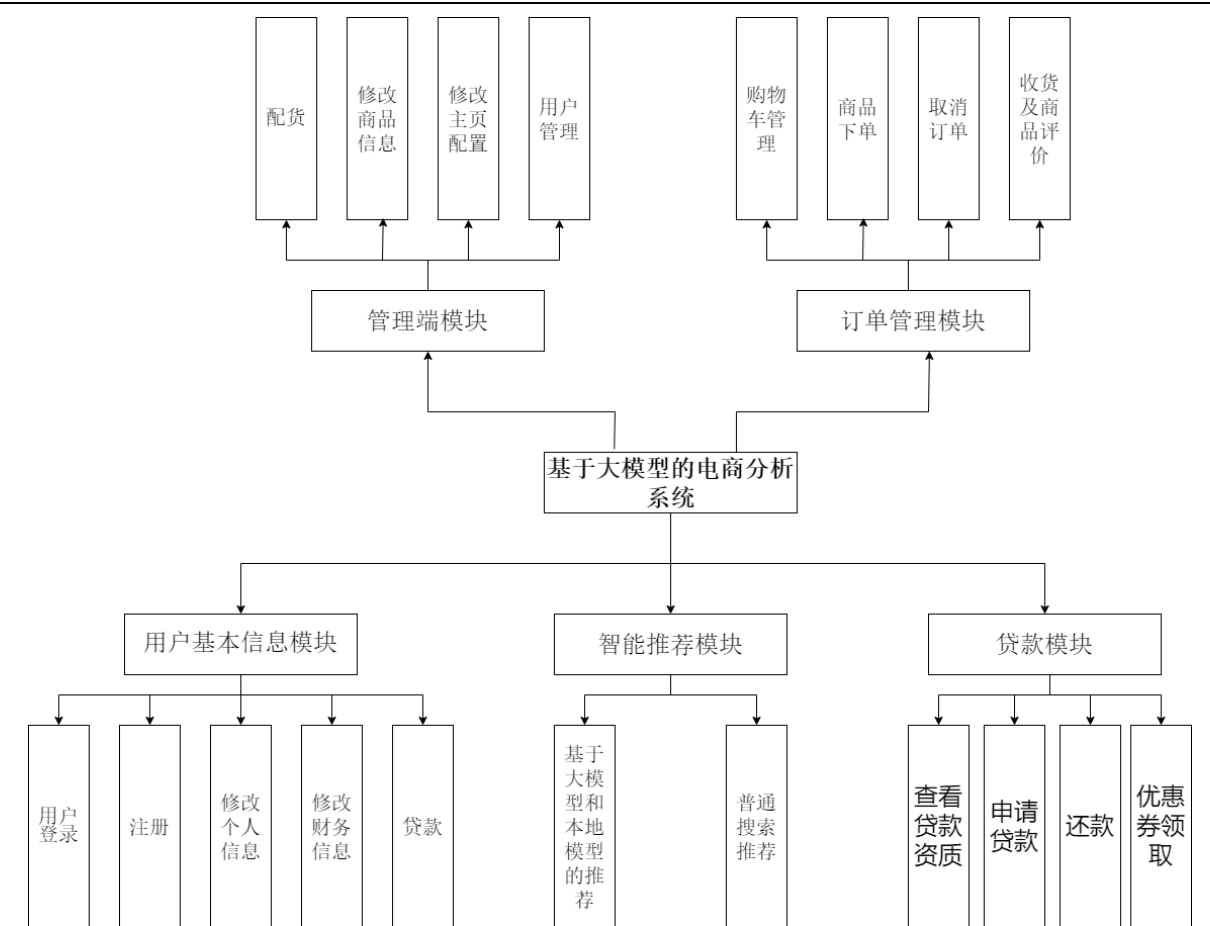


图 3-6 系统模块示意图

3.3.3 用户基本信息模块

该模块维护用户的基本信息，支持用户进行注册、登录以及浏览商品。用户可以通过简洁直观的界面添加商品到购物车、下单以及查看订单状态。此外，用户还能接收到基于个性化算法生成的商品推荐，这些推荐会显示在用户界面的显著位置，以提高用户的购物体验和销售效率。

3.3.4 管理端模块

为了确保电商平台的灵活性和高效性，管理端模块为商城管理员提供了一套全面的管理工具。管理员通过这个模块可以高效地编辑商品信息、更新商城公告以及管理配送信息。这个模块的设计考虑到了操作的便捷性和信息管理的安全性，使得管理员能够轻松地维护平台内容，保证信息的准确性和时效性。

3.3.5 订单管理模块

订单管理模块是处理用户购物流程的核心，它支持订单的创建、支付、查看以及配送跟踪。该模块能够处理用户从添加商品到购物车到完成购买的整个过程，包括订单状态更新、支付验证以及配送安排。通过这个模块，用户可以实时跟踪订单状态，而管理员也可以高效地管理订单流程，确保用户满意度。

3.3.6 智能推荐模块

智能推荐模块是该系统的技术亮点，它基于 Pytorch 框架和 Transformer 架构开发的 Seq2label 大语言模型。该模型结合 Kmeans 聚类算法，通过深入分析和处理大量用户评论，自动提炼出准确反映商品特性的标签。利用这些标签，智能推荐系统能够为用户推荐最合适的商品。通过细致的 fine-tuning 和词嵌入技术的应用，该模块不仅提升了推荐的准确性，也极大地丰富了用户的购物体验。

3.3.7 贷款模块

贷款模块做到了与金融方向紧密结合，它利用用户的购买信息预测用户在商城的信用评分，再通过用户输入的个人财务信息，通过随机森林算法为用于预测一个贷款的金额和期限。用户在选择贷款后可以获得商城优惠券，刺激用户进一步在商城内消费。

3.4 Seq2Label 模型

本项目的核心模块是一个基于深度学习的推荐系统，名为 Seq2label，旨在通过深度学习技术高效地将用户评论转化为商品标签，从而提高商品推荐的准确性和用户体验。该模型结合了先进的编码器-解码器架构，通过对用户评论的深入分析和理解，自动生成与评论内容紧密相关的商品标签。

对于 Seq2label 模型的构建方面，由于相较于翻译、文本生成、问答式大语言模型等复杂的自然语言处理任务，输出两个简短的标签是相对比较简洁的，并且面对潜在的海量评论，需要一个资源消耗较低，相应时间极快的模型完成这个任务。因此，本研究在借鉴 Transformer 模型的自注意力机制的同时，选择了基于循环神经网络（RNN）和 Luong 注意力机制的架构，以期望结合两者的优点来设计一个既能捕捉评论文本中的关键信息，又能快速响应的模型。

Seq2label 模型的核心设计思想在于利用双向 GRU 来处理输入的评论文本，从而有效捕获文本中的上下文信息。双向 GRU 的运用确保了模型能够从两个方向理解文本，增强了对语义的捕捉能力。紧接着，通过 Luong 注意力机制，模型能够聚焦于文本中对生成标签最关键的信息，这样不仅提高了标签生成的准确性，也增强了模型处理复杂文本的能力。

模型训练完成后，能够接受任意两条领域内的评论作为输入，并输出与这些评论最相关的标签。这些标签是通过分析和总结评论内容自动生成的，能够反映评论中的关键信息和用户意图。通过 Kmeans 算法对标签对应的词向量进行聚类，并选取聚类中心作为商品的最终标签，进一步提升了标签的准确性和代表性。

此外，Seq2label 模型的训练和验证过程中，充分利用了 CUDA 技术来加速计算，以应对大规模数据集的处理需求。通过这一系列精心设计和优化的训练流程，Seq2label 模型在实验中展现出了优异的性能，有效地提升了推荐系统的准确率和用户满意度，证明了该模型在电商推荐系统中的应用价值和潜力。

3.5 数据库设计

3.5.1 E-R 图

如图 3-7 所示，通过实体集、属性和联系集的图形化表示，展示了数据库的结构。本系统包含了商品分类、商品信息、商品评价、订单、订单地址、订单项、购物车项以及用户等实体集，这些实体集通过一系列的联系集相互连接，形成了电商金融系统的数据模型基础。

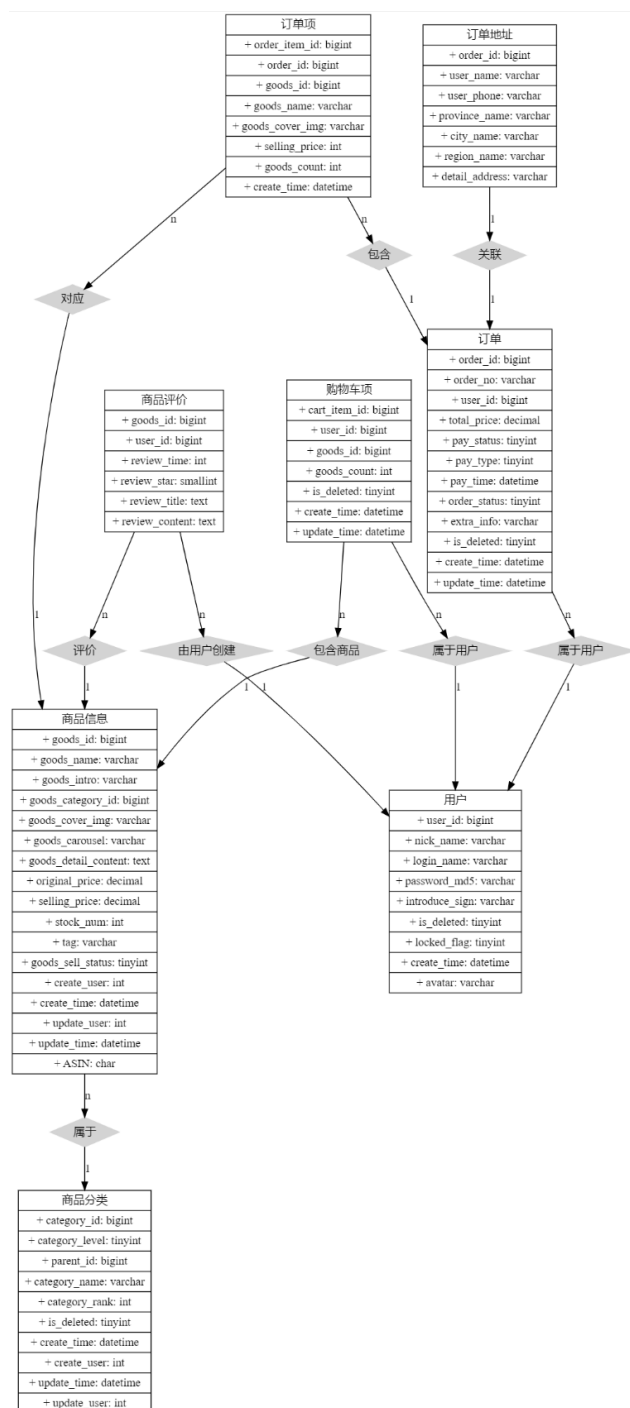


图 3-7 系统 E-R 图

3.5.2 实体集及其属性

商品分类实体集：分类 ID、分类级别、父分类 ID、分类名称、排序值、删除标识、创建时间、创建者 ID、修改时间、修改者 ID。其中分类 ID 是自增主键，分类名称是唯一键，父分类 ID 用于表示层级关系。

商品信息实体集：商品 ID、商品名称、商品简介、关联分类 ID、商品主图、商品轮播图、商品详情、原始价格、销售价格、库存数量、商品标签、商品销售状态、添加者 ID、添加时间、修改者 ID、修改时间、ASIN。其中商品 ID 是自增主键，添加时间为商品创建时间，商品主图和轮播图为商品图片存储于服务器中的地址。

用户实体集：用户 ID、昵称、登录名、密码、个性签名、注销标识、锁定标识、注册时间、头像。其中用户 ID 是自增主键，登录名是唯一键。

订单实体集：订单 ID、订单号、用户 ID、总价、支付状态、支付类型、支付时间、订单状态、额外信息、删除标识、创建时间、更新时间。其中订单 ID 是自增主键，订单号是唯一键。

订单地址实体集：订单 ID、收货人姓名、收货人手机号、省、市、区、详细地址。其中订单 ID 是主键，同时也是外键，关联到订单实体。

订单项实体集：订单项 ID、订单 ID、商品 ID、商品名称（快照）、商品主图（快照）、销售价格（快照）、商品数量（快照）、创建时间。其中订单项 ID 是自增主键，订单 ID 和商品 ID 为外键，分别关联到订单实体和商品信息实体。

购物车项实体集：购物车项 ID、用户 ID、商品 ID、商品数量、删除标识、创建时间、更新时间。其中购物车项 ID 是自增主键，用户 ID 和商品 ID 为外键，分别关联到用户实体和商品信息实体。

3.5.3 联系集及其属性

属于联系集（商品信息与商品分类）：一对多联系集，表示一个商品分类可以包含多个商品信息，但每个商品信息只属于一个商品分类。

评价联系集（商品评价与商品信息）：一对多联系集，表示一个商品信息可以有多个商品评价，但每个商品评价只能评价一个商品信息。

拥有联系集（订单与用户）：一对多联系集，表示一个用户可以拥有多个订单，但每个订单只属于一个用户。

关联联系集（订单地址与订单）：一对一联系集，表示每个订单有一个具体的订单地址。

包含联系集（订单项与订单）：一对多联系集，表示一个订单可以包含多个订单项，但每个订单项只属于一个订单。

对应联系集（订单项与商品信息）：一对一联系集，表示每个订单项对应一个具体的商品信息。

属于用户联系集（购物车项与用户）：一对多联系集，表示一个用户可以拥有多个购物车项，但每个购物车项只属于一个用户。

包含商品联系集（购物车项与商品信息）：一对一联系集，表示每个购物车项包含一个具体的商品信息

3.6 项目成本估算

估计项目成本如表 3.1 所示。

表 3.1 项目成本预算表			
序号	项目	工作量(天)	费用(元)
1	租用云服务器(1 年)	1	2000
2	调用 OpenAI 接口获取数据集	10	200
3	需求分析	5	100
4	深度学习框架调试	30	500
5	后端开发	20	800
6	前端开发	20	600
7	系统调试	3	400
8	测试	2	300
9	部署	1	200
	总计	81	4700

4 进度安排

进度安排如表 3.2 所示。

表 3.2 进度安排细则		
序号	各阶段工作	起止日期
1	选题并查阅相关文献，完成文献翻译和开题报告。	2023.12 - 2024.1
2	熟悉开发工具以及相关技术，包括深度学习框架和云服务器操作。	2024.1 - 2024.2
3	开发深度学习模块，包括算法研究和模型训练。	2024.2 - 2024.3
4	开发后端逻辑，包括商品信息管理、订单处理等功能。	2024.2 - 2024.3
5	开发前端界面和用户交互功能。	2024.3 - 2024.4
6	系统测试，包括单元测试、集成测试和性能测试。	2024.4 - 2024.5
7	系统调试和优化，确保系统稳定运行。	2024.5- 2024.6
8	撰写毕业论文，总结项目开发过程和成果。	2024.5 - 2024.6
9	提交毕设设计相关材料，制作论文答辩 PPT，进行毕业设计答辩。	2024.6

参考文献

- [1] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[J]. Advances in neural information processing systems, 2017, 30.
- [2] Radford A, Narasimhan K, Salimans T, et al. Improving language understanding by generative pre-training[J]. 2018.
- [3] Devlin J, Chang M W, Lee K, et al. Bert: Pre-training of deep bidirectional transformers for language understanding[J]. arXiv preprint arXiv:1810.04805, 2018.
- [4] 陆岷峰,高伦.大语言模型发展现状及其在金融领域的应用研究[J].金融科技时代,2023,31(08):32-38.
- [5] Sarwar B, Karypis G, Konstan J, et al. Analysis of recommendation algorithms for e-commerce[C]//Proceedings of the 2nd ACM Conference on Electronic Commerce. 2000: 158-167.
- [6] Driskill R, Riedl J. Recommender systems for e-commerce: Challenges and opportunities[C]//Proceedings of the AAAI-99 Workshop on AI for Electronic Commerce, USA. 1998.
- [7] Strubell E, Ganesh A, McCallum A. Energy and policy considerations for deep learning in NLP[J]. arXiv preprint arXiv:1906.02243, 2019.
- [8] 杨泉.嵌入式词向量的实现原理研究[J].计算机与数字工程,2023,51(11):2602-2607+2614.
- [9] Mikolov T, Chen K, Corrado G, et al. Efficient estimation of word representations in vector space[J]. arXiv preprint arXiv:1301.3781, 2013.
- [10] Pennington J, Socher R, Manning C D. Glove: Global vectors for word representation[C]//Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP). 2014: 1532-1543.
- [11] Moschitti A, Pang B, Daelemans W. Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)[C]//Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP). 2014.
- [12] Joulin A, Grave E, Bojanowski P, et al. Bag of tricks for efficient text classification[J]. arXiv preprint arXiv:1607.01759, 2016.
- [13] Cho K, Merriënboer V B, Gülçehre Ç, et al. Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation. [J]. CoRR, 2014, abs/1406. 1078.
- [14] Sutskever I, Vinyals O, Le Q V. Sequence to sequence learning with neural networks[J]. Advances in neural information processing systems, 2014, 27.
- [15] Luong M T, Pham H, Manning C D. Effective approaches to attention-based neural machine translation[J]. arXiv preprint arXiv:1508.04025, 2015.
- [16] Howard J, Ruder S. Universal language model fine-tuning for text classification[J]. arXiv preprint arXiv:1801.06146, 2018.
- [17] Ruder S. Neural transfer learning for natural language processing[D]. NUI Galway, 2019.
- [18] Phang J, Févry T, Bowman S R. Sentence encoders on stilts: Supplementary training on

- intermediate labeled-data tasks[J]. arXiv preprint arXiv:1811.01088, 2018.
- [19] MacQueen J. Some methods for classification and analysis of multivariate observations[C]//Proceedings of the fifth Berkeley symposium on mathematical statistics and probability. 1967, 1(14): 281-297.
- [20] 吴夙慧,成颖,郑彦宁等.K-means 算法研究综述[J].现代图书情报技术,2011,(05):28-35.
- [21] Zhang T, Ramakrishnan R, Livny M. BIRCH: an efficient data clustering method for very large databases[J]. ACM sigmod record, 1996, 25(2): 103-114.
- [22] Ester M, Kriegel H P, Sander J, et al. A density-based algorithm for discovering clusters in large spatial databases with noise[C]//kdd. 1996, 96(34): 226-231.
- [23] Breiman, L. (2001). Random Forests. Machine Learning, 45(1), 5-32.
- [24] Chen, X., Liaw, A., & Breiman, L. (2004). Using Random Forest to Learn Imbalanced Data. University of California, Berkeley.
- [25] Ni J, Li J, McAuley J. Justifying recommendations using distantly-labeled reviews and fine-grained aspects[C]//Proceedings of the 2019 conference on empirical methods in natural language processing and the 9th international joint conference on natural language processing (EMNLP-IJCNLP). 2019: 188-197.